AMENDMENTS TO THE CLAIMS

1. (currently amended) A method of preparing normalized and/or subtracted cDNAs characterized by comprising the steps of:

I) preparing uncloned full-length or full-coding length cDNAs ~~(testers)~~;

II) preparing polynucleotides ~~(drivers)~~ for normalization and/or subtraction;

III) conducting normalization and/or subtraction and removing tester/driver hybrids and non-hybridized polynucleotide drivers; and

IV) recovering the normalized and/or subtracted full-length or full-coding length cDNA.

2. (currently amended) The method of claim 1, wherein the cDNA tester of step I) is a reverse transcript of mRNA in the form of uncloned cDNA.

3. (currently amended) The method of claim 1 wherein said cDNA tester is ~~single strand~~single-stranded.

4. (previously presented) The method of claim 1, wherein in step III), normalization is conducted first, followed by subtraction.

5. (previously presented) The method of claim 1, wherein in step III), subtraction is conducted first, followed by normalization.

6. (previously presented) The method of claim 1, wherein in step III), said tester and normalization and subtraction drivers are mixed together and normalization and subtraction are conducted as a single step.

7. (canceled)

8. (currently amended) The method of claim 1, wherein step III) comprises the addition of an enzyme capable of cleaving ~~single-strand~~single-stranded RNA driver nonspecifically bound to ~~single-strand~~single-stranded cDNA and the cleaved ~~single-strand~~single-stranded RNA driver is removed.

9. (original) The method of claim 8 wherein said enzyme is single-strand-specific RNA endonuclease.

10. (original) The method of claim 8 wherein said enzyme is either selected from the group consisting of RNase I, RNaseA, RNase4, RNaseT1, RNaseT2, RNase2, and RNase3, or comprises a mixture thereof.

11. (original) The method of claim 8 wherein said enzyme is RNase I.

12. (previously presented) The method of claim 1, wherein said cDNA tester is prepared by CAP-trapping the 5' end of RNA.

13. (currently amended) The method of claim 1, wherein the preparation of said full-length or full-coding length cDNA tester comprises the following steps:

(1) synthesizing first strand cDNA by means of reverse transcriptase forming mRNA/cDNA hybrids;

(2) chemically binding a tag molecule to the diol structure of the 5' CAP($^{7Me}G_{ppp}N$) site of mRNA forming hybrids;.

(3) trapping ~~long-strand, full-coding, and/or full-length~~full-length or full-coding length cDNA hybrids; and

(4) removing ~~single-strand~~single-stranded mRNA ~~through~~ by digestion with an enzyme capable of cleaving ~~single-strand~~single-stranded mRNA.

14. (original) The method of claim 13 wherein said tag molecule is digoxigenin, biotin, avidin, or streptavidin.

15. (previously presented) The method of claim 1, wherein said polynucleotide driver for normalization and/or subtraction is RNA and/or DNA.

16. (original) The method of claim 15, wherein said DNA driver is cDNA.

17. (currently amended) The method of claim 1, wherein said normalization driver comprises cellular mRNA from the same library, from the same tissue, or the same cDNA population as ~~what is~~the cDNA to be normalized.

18. (currently amended) The method of claim 1, wherein said normalization driver comprises ~~single-strand~~single-stranded cDNA obtained from the same library, the same tissue, or the same cDNA population as ~~what is~~the cDNA to be normalized.

19. (currently amended) The method of claim 1, wherein said subtraction driver comprises cellular mRNA from a library, tissue, or cDNA population differing from ~~what is~~ the cDNA to be subtracted.

20. (currently amended) The method of claim 1, wherein said subtraction driver comprises ~~single strand~~single-stranded cDNA from a library, tissue, or cDNA population differing from ~~what is~~the cDNA to be normalized.

21. (currently amended) The method of claim 1, further comprising a step V) of preparing a ~~second~~ complementary strand of the recovered cDNA and ~~performing~~ cloning the resulting double-stranded cDNA.

22. (currently amended) ~~:~~ A method of preparing normalized and/or subtracted full-length or full-coding length cDNAs ~~characterized by~~ comprising the steps of:

I) preparing cDNAs ~~(~~testers~~)~~ not cloned in a plasmid;

II) preparing polynucleotides ~~(~~drivers~~)~~ for normalization and/or subtraction;

III) conducting normalization and/or subtraction and removing tester/driver hybrids and non-hybridized polynucleotide drivers; and

IV) recovering the normalized and/or subtracted full-length or full-coding length cDNA.

23. (original) The method of claim 22, wherein in step III), normalization is conducted first, followed by subtraction.

24. (original) The method of claim 22, wherein in step III), subtraction is conducted first, followed by normalization.

25. (original) The method of claim 22, wherein in step III), said tester and normalization and subtraction drivers are mixed together and normalization and subtraction are conducted as a single step.

26. (canceled.)

27. (currently amended) The method of claim 22, wherein step III) comprises the addition of an enzyme ~~capable of cleaving~~that cleaves single-strand<u>ed</u> RNA driver nonspecifically bound to ~~single-strand~~<u>single-stranded</u> cDNA and the cleaved ~~single-strand~~<u>single-stranded</u> RNA driver is removed.

28. (original) The method of claim 27, wherein said enzyme is single-strand-specific RNA endonuclease.

29. (original) The method of claim 27, wherein said enzyme is either selected from the group consisting of RNase I, RNaseA, RNase4, RNaseT1, RNaseT2, RNase2, and RNase3, or comprises a mixture thereof.

30. (original) The method of claim 27, wherein said enzyme is RNase I.

31. (previously presented) The method of claim 22, wherein said cDNA tester is prepared by CAP-trapping the 5' end of RNA.

32. (currently amended) The method of claim 22, wherein said normalization driver comprises cellular mRNA from the same library, the same tissue, or the same cDNA population ~~as what~~ <u>the cDNA</u> is to be normalized.

33. (currently amended) The method of claim 22, wherein said normalization driver comprises ~~single-strand~~single-stranded cDNA obtained from the same library, the same tissue, or the same cDNA population as ~~what is~~ the cDNA to be normalized.

34. (currently amended) The method of claim 22, wherein said subtraction driver comprises cellular mRNA from a library, tissue, or cDNA population differing from ~~what is~~ the cDNA to be subtracted.

35. (currently amended) The method of claim 22, wherein said subtraction driver comprises ~~single-strand~~single-stranded cDNA from a library, tissue, or cDNA population differing from ~~what is~~the cDNA to be normalized.

36. (currently amended) The method of claim 22, further comprising a step V) of preparing a ~~second~~ complementary strand of the recovered full-length or full-coding length cDNA and ~~performing~~ cloning the resulting double-stranded full-length or full-coding length cDNA.

37. (currently amended) A method of preparing normalized and subtracted full-length or full-coding length cDNA comprising the steps of:

I) preparing cDNA~~s (~~tester~~)~~;

II) preparing polynucleotide~~s (~~drivers~~)~~ for normalization and subtraction;

III) conducting the normalization and subtraction as a single step by mixing together the tester and the drivers; and

IV) recovering the normalized and subtracted full-length or full-coding length cDNA.

38. (original) The method of claim 37, wherein the cDNA tester is cloned or uncloned cDNA.

39. (currently amended) The method of claim 37, wherein the cDNA tester is the reverse transcript of mRNA in the form of uncloned cDNA.

40. (currently amended) The method of claim 37, wherein the cDNA tester is ~~single strand~~single-stranded.

41. (canceled.)

42. (currently amended) The method of claim 37, wherein step III) comprises the addition of an enzyme capable of cleaving single-strand RNA driver nonspecifically bound to ~~single strand~~single-stranded cDNA and the cleaved ~~single strand~~single-stranded RNA driver is removed.

43. (original) The method of claim 42, wherein said enzyme is single-strand-specific RNA endonuclease.

44. (original) The method of claim 42, wherein said enzyme is either selected from the group consisting of RNase I, RNaseA, RNase4, RNaseT1, RNaseT2, RNase2, and RNase3, or comprises a mixture thereof.

45. (original) The method of claim 42, wherein said enzyme is RNase I.

46. (previously presented) The method of claim 37, wherein said cDNA tester is prepared by CAP-trapping 5' end of RNA.

47. (currently amended) The method of claim 37, wherein the preparation of said full-length or full-coding length cDNA tester comprises the following steps:

(1) synthesizing first strand cDNA by means of reverse transcriptase forming

mRNA/cDNA hybrids;

(2) chemically binding a tag molecule to the diol structure of the 5' CAP($^{7Me}G_{ppp}N$) site of mRNA forming hybrids;

(3) trapping ~~long strand, full coding, and/or full length~~full-length or full-coding length cDNA hybrids; and

(4) removing ~~single strand~~single-stranded mRNA ~~through~~ by digestion with an enzyme ~~capable of cleaving single strand~~that cleaves single-stranded mRNA.

48. (original) The method of claim 47, wherein said tag molecule is digoxigenin, biotin, avidin, or streptavidin.

49. (previously presented) The method of claim 37, wherein said polynucleotide driver for normalization and/or subtraction is RNA and/or DNA.

50. (previously presented) The method of claim 49, wherein said DNA driver is cDNA.

51. (currently amended) The method of claim 37, wherein said normalization driver comprises cellular mRNA from the same library, the same tissue, or the same cDNA population as ~~what is~~ the cDNA to be normalized.

52. (currently amended) The method of claim 37, wherein said normalization driver comprises ~~single strand~~single-stranded cDNA obtained from the same library, the same tissue, or the same cDNA population as ~~what is~~the cDNA to be normalized.

53. (currently amended)The method of claim 37, wherein said subtraction driver comprises cellular mRNA from a library, tissue, or cDNA population differing from ~~what is~~the cDNA to be subtracted.

54. (currently amended) The method of claim 37, wherein said subtraction driver comprises ~~single-strand~~single-stranded cDNA from a library, tissue, or cDNA population differing from ~~what is~~the cDNA to be normalized.

55. (currently amended) The method of claim 37, further comprising a step V) of preparing a ~~second~~ complementary strand of the recovered full-length or full-coding length cDNA and ~~performing~~ cloning the resulting double-stranded full-length or full-coding length cDNA.

56. (currently amended) A method of preparing normalized and/or subtracted full-length or full-coding length cDNA comprising the steps of:

(a) preparing cDNA ~~(~~tester~~)~~;

(b) preparing normalization and/or subtraction RNA ~~(~~driver~~)~~;

(c) conducting normalization and/or subtraction in two steps in any order, or

conducting normalization/subtraction as a single step and mixing the normalization/subtraction RNA driver with said cDNA tester;

(d) adding an enzyme ~~capable of cleaving single strand~~that cleaves single-stranded sites on RNA drivers non-

specifically bound to cDNA tester;

(e) removing said ~~single strand~~single-stranded RNA driver cleaved in step d) from the tester and

removing tester/driver hybrids; and

(f) recovering the normalized and/or subtracted full-length or full-coding length cDNA.

57. (original) The method of claim 56, wherein the cDNA tester is cloned or uncloned cDNA.

58. (currently amended) The method of claim 56, wherein the cDNA tester is a reverse transcript of mRNA in the form of uncloned cDNA.

59. (currently amended) The method of claim 56, wherein said cDNA tester is ~~single strand~~single-stranded.

60. (previously presented) The method of claim 56, wherein in step c), normalization is conducted first, followed by subtraction.

61. (previously presented) The method of claim 56, wherein in step c), subtraction is conducted first, followed by normalization.

62. (previously presented) The method of claim 56, wherein in step c), said tester and normalization and subtraction drivers are mixed together and normalization and subtraction are conducted as a single step.

63. (currently amended) The method of claim 56, wherein said normalized and/or subtracted cDNA is ~~long-strand, full-coding, and/or full-length~~full-length or full-coding length cDNA.

64. (previously presented) The method of claim 56, wherein the enzyme of said step d) is either selected from the group consisting of RNase I, RNaseA, RNase4, RNaseT1, RNaseT2, RNase2, and RNase3, or comprises a mixture thereof.

65. (previously presented) The method of claim 56, wherein the enzyme of said step d) is RNase I.

66. (previously presented) The method of claim 56, wherein said cDNA tester is prepared by CAP-trapping the 5' end of RNA.

67. (currently amended) The method of claim 56, further comprising the step g) of preparing a ~~second~~ complementary strand of the recovered cDNA and ~~performing~~ cloning the resulting double-stranded cDNA.

68. (previously presented) The method of claim 1, wherein said tester/driver hybrids are bound to tag molecules.

69. (original) The method of claim 68, wherein said tag molecule is avidin, streptavidin, biotin, digoxigenin, an antibody, or an antigen.

70. (previously presented) The method of claim 1, wherein said tester/driver hybrids are removed through the use of a matrix.

71. (original) The method of claim 70, wherein said matrix is comprised of magnetic beads or agarose beads.

72. (currently amended) The method of claim 71, wherein said magnetic beads or agarose beads are covered by or bound to ~~any~~ a tag molecule ~~capable of binding~~that binds to a tag ~~molecules~~ molecule bound to a tester/driver ~~hybrids~~hybrid.

73. (currently amended) The method of claim 71, wherein said magnetic beads or agarose beads are covered by or bound to a tag molecule ~~capable of binding~~that binds to avidin, streptavidin, biotin, digoxigenin, an antibody, or an antigen bound to a tester/driver hybrid.

74. (currently amended) The method of claim 72, wherein said ~~antibody~~ tag molecule covering said beads or ~~said antibody binding~~ bound to said beads is an antiantigen antibody, antibiotin antibody, antiavidin antibody, antistreptavidin antibody, or antidigoxigenin antibody.

75. (currently amended) The method of claim 1, wherein said tester/driver hybrid is removed ~~through the use of~~ using streptavidin/phenol.

76. (currently amended) The method of claim 1, wherein ~~hydroxyapadite~~ hydroxyapatite and nonlabeled RNA are employed to remove said tester/driver hybrid.

77. (currently amended) A method of removing RNA nonspecifically bound to DNA ~~by~~ comprising processing nonspecifically bound RNA/DNA hybrids with an enzyme ~~capable of degrading~~ that degrades ~~single strand~~single-stranded RNA.

78. (original) The method of claim 77, wherein said enzyme is either selected from the group consisting of RNase I, RNaseA, RNase4, RNaseT1, RNaseT2, RNase2, and RNase3, or comprises a mixture thereof.

79. (original) The method of claim 77, wherein said enzyme is RNase I.

80. (previously presented) The method of claim 77, wherein said RNA/DNA hybrid is a product of normalization.

81. (previously presented) The method of claim 77, wherein said RNA/DNA hybrid is a product of subtraction.

82. (previously presented) The method of claim 77, wherein said RNA/DNA hybrid is the product of a method comprising the steps of normalization and subtraction in any order or of a method comprising a single normalization/subtraction step.

83. (currently amended) A method of isolating ~~single strand~~single-stranded cDNA comprising the steps of treating a hybrid comprising RNA nonspecifically bound to cDNA with an enzyme

~~capable of degrading~~ that degrades ~~single strand~~single-stranded RNA, removing the degraded ~~single strand~~single-stranded RNA, and recovering the cDNA.

84. (currently amended) A method of preparing normalized and/or subtracted cDNA comprising the steps of adding an enzyme ~~capable of degrading single strand~~that degrades single-stranded RNA driver nonspecifically bound to cDNA tester, and removing the degraded ~~single strand~~single-stranded RNA driver.

85. (currently amended) The method of claim 77, wherein said DNA or cDNA is ~~long chain, full coding, and/or full length~~full-length or full-coding length cDNA.

86. (previously presented) The method of claim 1 employed to prepare one, two, or more libraries.

87. (canceled.)

88. (canceled.)

89. (new) The method of claim 1, in which subtraction is performed and normalization is performed to a R_0T value of from 5 to 10.

90. (new) The method of claim 13, wherein the chemical tagging is performed on ice.

REMARKS

The Office Action of April 2, 2003 presents the examination of claims 1-86, claims 87 and 88 being deemed withdrawn from consideration. Claims 87 and 88 are canceled herein; Applicant reserves the right to file an application directed to the canceled subject matter pursuant to 35 USC § 120.

The present paper also adds new claims 89 and 90. Support for claim 89 with respect to R_0T values is found in Table 2 at page 40. Support for claim 90 is found, e.g. at page 8, in the description of Figure 6, taken with the disclosure at page 36, under "Biotinylation of RNA".

Many claims are amended to provide somewhat clearer language to the recitations.

Rejections under 35 U.S.C. § 112, second paragraph

Claims 1-36, 74 and 76 stand rejected under 35 U.S.C. § 112, second paragraph as being indefinite for failure to distinctly claim the subject matter of the invention. Applicant has amended the claims so as to obviate this rejection.

Specifically, the term "characterized by" has been deleted as suggested by the Examiner. The term "said antibody" has been corrected to "said tag" in claim 74 to provide proper antecedent basis. In claim 76, the word "hydroxyapatite" has been correctly spelled.

Anticipation rejections

Claims 1-4, 6-23, 25-73 and 77-86 stand rejected under 35 U.S.C. § 102(a) as anticipated by Carninci et al. (2000). This rejection is respectfully traversed. Reconsideration and withdrawal thereof are requested.

The Carninci (2000) reference was published in the time between the filing date of the priority application. Applicant is preparing a verified English translation of the priority application JP 2000-255402 and will timely file it to overcome the instant rejection.

Claims 1-3, 6, 7, 15-22, 25, 26, 32-41 and 49-55 stand rejected under 35 U.S.C. § 102(e) as anticipated by Chang '874. This rejection is respectfully traversed. Reconsideration and withdrawal thereof are requested.

The invention as recited in claims 1-3, 6, 7, 15-22 25, 26 32-41 and 49-55 is directed to full-length or full-coding length cDNA and libraries of such full-length or full-coding length cDNA. In contrast Chang discloses cDNAs and libraries thereof that do not meet this limitation. Support for the recitation in claims 1, 22, etc. of "full-length" or "full-coding" length cDNA or libraries thereof is provided by the specification at, e.g. page 5, in paragraph 18.

Chang discloses the preparation of cDNAs that are not full-length or full-coding length, a subtraction step, then isolation of clones and finally a screening of the (not full-length or full-coding length) cDNA library in order to discover the presence of one or more full-length or full-coding length cDNAs. (see Example 1, from column 23, line 51 to the end of column 25.) That Chang fails to enable the preparation of full-length cDNA is even admitted by the Examiner at page 9, lines 8-10.

Chang provides no motivation to prepare full-length or full-coding length cDNA libraries for subtraction as he is interested in finding of only one, specific clone and carries out a further step of screening and search for this specific clone. This is evidenced by the fact that Chang prepares cDNAs using photo-biotinylation. However, photo-biotinylation degrades cDNAs (see Fargnoli et al., *Analytical Biochemistry*, 187:364-373 (1990), at page 365, at the

bottom of the page, to follow later as Exhbit 1). As a result, photo-biotinylation does not allow an efficient preparation of full-length or full-coding length cDNAs.

cDNAs or cDNA libraries prepared according to Chang are not full-length or full-coding length cDNA or cDNA libraries. The presence of few or some full-length or full-coding length cDNAs does not render the cDNA preparation or a cDNA library a "full-length" preparation or library. With respect to definition of full-length libraries, the Examiner might refer to the paper of Marra et al. (*Nature Genetics* (1999), attached as Exhibit 2. In Figure 2 of the Marra et al. paper, the definition of full-length libraries versus EST (normal) libraries is demonstrated as a graphical explanation of the approximate full-length ratio.

Furthermore, to obtain any full-length or full-coding length cDNA clone, Chang must perform a step of screening, sequencing and homology database searching, with merely a hope of finding a full-length or full-coding length clone. Chang clearly indicates at column 25, lines 36-37 that they "found" a full-length clone from a not-full-length cDNA library. This illustrates the mere hope or the role of pure luck in finding a specifically searched clone using the method of Chang. A library prepared according to Chang might not necessarily and always include a desired full-length or full-coding length clone.

The steps of screening, sequencing and homology searching with the aim of finding a particular full-length or full-coding length clone from subtraction libraries were inconveniences in the art to be overcome at the time the invention was made. *See*, for example, Sagerstrom et al., *Annu. Rev, Biochem.*, 66:751-83 (1997) (of record), from page 777, last paragraph to page 778: "Two major problems with current subtractive or positive selection techniques are (a) an inability to easily isolate full-length clones after subtraction ...".

In conclusion, Chang does not disclose, and does not enable, all the features of the invention recited in claim 1 as he does not prepare full-length or full-coding length tester cDNA before subtraction and does not obtain a subtracted complete full-length or full-coding length cDNA preparation. Accordingly, the instant rejection should be withdrawn.

Claims 1-7, 15-18, 21-26, 32, 33, 36-41, 50-52 and 55 are rejected under 35 U.S.C. § 102(b) as being anticipated by Ruppert et al. '637. This rejection is respectfully traversed. Reconsideration and withdrawal thereof are requested.

Applicant submits that the invention as described in the rejected claims is not described by Ruppert. In particular, claim 1 and the other independent claims 22 and 37, recite a process comprising preparing cDNA tester and polynucleotide (e.g., RNAs) drivers.

On the contrary Ruppert discloses the preparation of mRNA tester and cDNA drivers, i.e. the opposite of the invention. See Fig. 1 of Ruppert. This difference is quite important. Claim 1 relates to the preparation of cDNA tester which can be used for normalization and/or subtraction. On the contrary, the mRNA tester disclosed by Ruppert can be efficiently used for normalization, but not for subtraction.

Ruppert's method only discloses normalization but NOT subtraction. Ruppert, unfortunately, improperly used the term "subtraction" to mean "self-subtraction", i.e. "normalization". The only method disclosed in Ruppert is the selection of low abundant mRNA (column 6, line 33-38). This is a normalization method but not a subtraction method. The Examiner should also see Example 1, column 9. The method disclosed by Ruppert is ONLY normalization, not subtraction, even if the word "subtraction" is used. Applicant supposes that

the use of the term "subtraction" by Ruppert, though this is not what is actually done in his method, is the cause of the Examiner's misinterpretation of the reference.

As explained above, mRNA tester (disclosed by Ruppert) cannot be efficiently used in subtraction. This is because mRNA is subject to degradation. In fact, while the hybridization during normalization process is carried out to the typical RoT value of 1-10, the time for carrying out the subtraction is longer, typically around RoT 10 to 500. This means that if mRNA is used as tester (like in Ruppert), this tester (mRNA) will be subjected to degradation and will not be efficiently recovered. In particular, full-length or full-coding length RNA testers cannot be efficiently recovered. The problem of degradation of mRNA is noted by Ruppert; *see* for example column 5, at lines 56-58.

As Ruppert does not disclose all of the features of the invention as recited in claims 1-7, 15-18, 21-26, 32, 33, 36-41, 50-52 and 55, the reference does not anticipate these claims and the instant rejection should be withdrawn.

Claims 77-86 stand rejected under 35 U.S.C. § 102(b) as anticipated by Carninci et al. (1996). This rejection is respectfully traversed. Reconsideration and withdrawal thereof are requested.

The Examiner asserts that Carninci et al. (1996) discloses the digestion of non-specifically bound hybrids by RNAse I, citing the disclosure at column 2, and part C of Figure 1. Careful examination of this disclosure shows that Carninci et al. (1996) in fact discloses use of RNAse I to digest the single-stranded portion of RNAs which are not completely protected by full-length first-strand cDNA synthesis. Such RNA-DNA hybrids are not "non-specific" hybrids

and so the reference fails to disclose each feature of the invention recited in claims 77-86. Accordingly, the instant rejection should be withdrawn.

Obviousness rejections

Claims 4, 5, 8-14, 23, 24, 27-31, 42-48, 56-73 and 77-86 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over Chang '874 in view of Carninci et al. (1996). This rejection is respectfully traversed. Reconsideration and withdrawal thereof are requested.

Applicant submits that the Examiner fails to establish proper *prima facie* obviousness of the claimed invention. In particular, the combination of the cited references fails to provide each and every feature of the claimed invention. Also, there is not any motivation to modify the Chang reference as the Examiner alleges to obtain the present invention.

The Examiner alleges that Chang '874 indicates the desirability of removing non-specific hybrids between RNA and DNA formed during normalization or subtraction and cites Carninci et al. (1996) for alleged disclosure of use of RNAse I to perform such a reaction.

The Examiner's characterization of the Chang reference is incorrect. Applicant notes that the portion of Chang '874 pointed to by the Examiner as disclosing the desirability of removing RNA-DNA hybrids during normalization or subtraction does not disclose such. Rather, Chang describes there the desirability of removing non-full-length cDNA-RNA hybrids following first-strand cDNA synthesis after a capping reaction. This is disclosure identical to that provided by Carninci (1996) as explained above. Thus, Chang '874 does not provide any motivation to remove non-specific RNA-DNA hybrids during any normalization or subtraction step. Furthermore, combining Chang '874 with Carninci et al. (1996) as suggested by the Examiner does not provide any disclosure of such a step. Still further, the use by Chang of photo-

biotinylation precludes preparation of any full-length or full-coding length cDNA library, and renders combination with Carninci (1996) inconsistent.

For all of the above reasons, the combined references fail to disclose a recited feature of the invention. Accordingly, the instant rejection fails and should be withdrawn.

Claim 74 is rejected under 35 U.S.C. § 103(a) as being unpatentable over Chang '874 in view of Carninci et al. (1996) in further view of Bouma. This rejection is respectfully traversed. Reconsideration and withdrawal thereof are requested.

Claim 74 depends ultimately from claim 1 and recites a further feature related to a specific method for removing tester/driver hybrids from the reaction.

The failure of the combination of Chang and Carninci et al. to describe even the basic invention of claim 1 is explained above. For example, the use by Chang of photo-biotinylation precludes preparation of any full-length or full-coding length cDNA library, and renders combination with Carninci (1996) inconsistent. Bouma is described by the Examiner as disclosing avidin and anti-biotin antibodies can be used equivalently to capture biotinylated moieties. Bouma does not cure the deficiencies of Chang and Carninci et al. as to the failure of the combined references to describe every feature of the claimed invention, i.e. the limitations recited in claim 1, nor does Bouma cure the inconsistency between the methods of Chang and the methods of Carninci et al. (1996) that preclude their combination as suggested by the Examiner. Accordingly, the instant rejection fails and should be withdrawn.

Claim 75 is rejected under 35 U.S.C. § 103(a) as being unpatentable over Ruppert et al. '637 in view of Mishra et al. This rejection is respectfully traversed. Reconsideration and withdrawal thereof are requested.

The Examiner fails to establish *prima facie* obviousness of the claimed invention. Claim 75 depends from claim 1, but adds the further recitation that hybrids of tester and driver are removed by streptavidin/phenol extraction. Mishra is cited as teaching this latter recitation.

The failure of Ruppert et al. '637 to describe even the basic invention of claim 1 is explained above. Mishra does not remedy this failure. For example, Mishra does not disclose or suggest that a cDNA tester should be used with a mRNA driver. Thus, the combination of the references suggested by the Examiner fails to teach or suggest every recitation of the claimed invention and so fails to establish *prima facie* obviousness of the invention. Accordingly, the instant rejection should be withdrawn.

Claim 76 is rejected under 35 U.S.C. § 103(a) as being unpatentable over Chang '874 in view of Carninci et al. (1996) and Lavery '548. This rejection is respectfully traversed. Reconsideration and withdrawal thereof are requested.

Claim 76 depends ultimately from claim 1, and further recites features related to the removal of tester/driver hybrids by hydroxyapatite. The disclosures of Chang and Carninci, and the failure of the combination of these two references to describe the basic invention of claim 1, is explained above.

Lavery is cited by the Examiner as disclosing use of hydroxyapatite as equivalent to biotin-streptavidin for capture of desired nucleic acids. Notwithstanding that in the present invention the hydroxyapatite is used to remove undesired nucleic acids, adding Lavery to the

combination of Chang '874 and Carninci et al. (1996) fails to remedy the deficiencies of Chang and Carninci to describe or suggest the basic invention. Thus, the further combination also fails to describe or suggest the basic invention, the Examiner fails to establish *prima facie* obviousness of the invention and the instant rejection fails and should be withdrawn.

Applicant submits that the present application well describes and claims patentable subject matter. The favorable action of withdrawal of the standing rejections and allowance of the application is respectfully requested.

Should there be any outstanding matters that need to be resolved in the present application, the Examiner is respectfully requested to contact Mark J. Nuell (Reg. No. 36,623) at the telephone number of the undersigned below, to conduct an interview in an effort to expedite prosecution in connection with the present application.

Pursuant to the provisions of 37 C.F.R. §§ 1.17 and 1.136(a), Applicant respectfully petitions for a three (3) month extension of time for filing a response in connection with the present application. The required fee of $950.00 is attached hereto.

If necessary, the Commissioner is hereby authorized in this, concurrent, and future replies, to charge payment or credit any overpayment to Deposit Account No. 02-2448 for any additional fees required under 37 C.F.R. § 1.16 or under 37 C.F.R. § 1.17; particularly, extension of time fees.

Respectfully submitted,

BIRCH, STEWART, KOLASCH & BIRCH, LLP

By _____
Mark J. Nuell, #36,623

P.O. Box 747
Falls Church, VA 22040-0747
(703) 205-8000

DRN/mua
2870-0173P

Attachment: Exhibit 2, Marra et al.

letter

An encyclopedia of mouse genes

Marco Marra[1], LaDeana Hillier[1], Tamara Kucaba[1], Melissa Allen[1], Robert Barstead[2], Catherine Beck[2], Angela Blistain[1], Maria Bonaldo[3], Yvette Bowers[1], Louise Bowles[1], Marco Cardenas[1], Ann Chamberlain[1], Julie Chappell[1], Sandra Clifton[1], Anthony Favello[1], Steve Geisel[1], Marilyn Gibbons[1], Njata Harvey[1], Francesca Hill[4], Yolanda Jackson[1], Sophie Kohn[1], Greg Lennon[4,5], Elaine Mardis[1], John Martin[1], LeeAnne Mila[4], Rhonda McCann[1], Richard Morales[1], Deana Pape[1], Barry Person[1], Christa Prange[4], Erika Ritter[1], Marcelo Soares[3], Rebecca Schurk[1], Tanya Shin[1], Michele Steptoe[1], Timothy Swaller[1], Brenda Theising[1], Karen Underwood[1], Todd Wylie[1], Tamara Yount[1], Richard Wilson[1] & Robert Waterston[1]

The laboratory mouse is the premier model system for studies of mammalian development due to the powerful classical genetic analysis[1] possible (see also the Jackson Laboratory web site, http://www.jax.org/) and the ever-expanding collection of molecular tools[2,3]. To enhance the utility of the mouse system, we initiated a program to generate a large database of expressed sequence tags (ESTs) that can provide rapid access to genes[4–16]. Of particular significance was the possibility that cDNA libraries could be prepared from very early stages of development, a situation unrealized in human EST projects[7,12]. We report here the development of a comprehensive database of ESTs for the mouse. The project, initiated in March 1996, has focused on 5′ end sequences from directionally cloned, oligo-dT primed cDNA libraries. As of 23 October 1998, 352,040 sequences had been generated, annotated and deposited in dbEST, where they comprised 93% of the total ESTs available for mouse. EST data are versatile and have been applied to gene identification[17], comparative sequence analysis[18,19], comparative gene mapping and candidate disease gene identification[20], genome sequence annotation[21,22], microarray development[23] and the development of gene-based map resources[24].

Our aims were to maximize gene discovery and to provide a broad overview of genes expressed throughout development. To these ends, more than one-half (178,500) of submitted ESTs were from 15 normalized libraries, which feature reduced redundancy[25], and more than one-third (124,679) were from 26 early-stage libraries (Table 1). Libraries from nine organs (heart, kidney, liver, lung, lymph node, placenta, spleen, thymus, uterus), smooth and striated muscle, blood cells, epithelial tissue, regions of the intestine, endocrine tissue, sex glands and whole embryos were sequenced. To increase the likelihood that ESTs would fall in regions of the cDNA coding for protein, most sequencing was performed from the 5′ end, but some 3′ ESTs were generated either intentionally, as for the Sugano libraries (Table 1), or indirectly, as a consequence of EST length exceeding cDNA insert size. Sequences from each library were monitored to assess library content, complexity and overall suitability for further sequencing. Not all libraries sequenced with the same success: sequence failures were categorized as technical, in which some aspect of the DNA purification or sequencing protocol was at fault, or non-technical, which encompassed sequences that were mitochondrial or bacterial in origin or were from non-recombinant clones. Libraries exhibiting higher frequencies of non-technical failures were considered low quality and were not sampled extensively. To assess library complexity, all ESTs from a library were compared routinely with each other ('clustering'). A high fraction of unique ESTs was taken as an indication of the increased complexity of the library; these were targeted preferentially for extensive sequencing.

ESTs are single-pass unedited sequences; hence, sequence data quality is of utmost importance. To measure the accuracy of the trimmed EST data, the automatic base calls generated by PHRED (refs 26,27) were compared with mouse coding sequences available from a database maintained at the National Center for Biotechnology Information (referred to here as the mouse mRNA set; G. Schuler, pers. comm.). Discrepancies and their positions in the ESTs were identified and categorized as base substitutions, deletions or insertions (Fig. 1). Discrepancies were not examined individually; thus, sequence polymorphisms, alternative splicing events or errors in the mouse coding sequences, although not resulting from faulty EST base calls, would be included in this analysis. Base substitutions were found most frequently, appearing at approximately twice the rate of insertions or deletions. All three types of discrepancies were most prevalent in the initial base pairs and showed decreasing frequencies as a function of EST length. These levels of accuracy, which represent increases over those previously reported[12], did not inhibit our analysis of ESTs by BLAST or other programs.

Library quality contributes substantially to the success of an EST project. As a measure of quality, we estimated the frequencies of inverted cDNA inserts by comparing ESTs with the mouse mRNA set. We identified 53,303 matches, which represented 84% of the sequences in the mouse mRNA set. Most matches (94%) were to the correct strand, although 6% matched the complement (wrong) strand. For two-thirds of the wrong-strand matches (4% of total matches), at least two ESTs mapped to the same position on the wrong strand, suggesting the match resulted from non-random events during library construction. Some fraction of these 'verified' wrong-strand matches may identify overlapping transcription units, although this was not tested. Thus, only 2% of the matches were wrong-strand single occurrences, possibly resulting from failures in directional cloning or human error.

[1]Washington University Genome Sequencing Center, 4444 Forest Park Boulevard, St. Louis, Missouri 63108, USA. [2]Oklahoma Medical Research Foundation, Program in Molecular & Cell Biology, 825 NE 13th Street, Oklahoma City, Oklahoma 73104, USA. [3]The University of Iowa, Unit 41, 451 Eckstein Medical Research Building, Iowa City, Iowa 52242, USA. [4]The I.M.A.G.E. Consortium, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, 7000 East Ave/L-452 Livermore, California 94550, USA. [5]GeneLogic, Inc. Genomics, 708 Quince Orchard Road, Gaithersburg, Maryland 20878, USA. Correspondence should be addressed to M.M. (e-mail: mmarra@alu.wustl.edu).

Table 1 • Summary of ESTs generated and submitted to dbEST

Library	Submitted	Attempted	Fraction submitted
Soares mouse embryo NbME13.514.5	35,541	46,908	0.758
Soares mouse mammary gland NbMMG	32,058	39,837	0.805
Soares 2NbMT	23,452	29,409	0.797
Soares mouse p3NMF19.5	21,648	27,785	0.779
Stratagene mouse skin (#937313)	15,553	20,773	0.749
Knowles-Solter mouse 2 cell	13,133	18,690	0.703
Barstead mouse myotubes MPLRB5	12,392	15,194	0.816
Soares mouse lymph node NbMLN	11,196	14,916	0.751
Knowles-Solter mouse blastocyst B1	10,896	17,339	0.628
Soares mouse 3NbMS	10,513	13,028	0.807
Soares mouse 3NME125	10,429	12,844	0.812
Stratagene mouse heart (#937316)	9,215	12,068	0.764
Barstead mouse irradiated colon MPLRB7	9,131	12,407	0.736
Soares mouse NML	8,971	10,966	0.818
Soares mouse NbMH	7,490	8,844	0.847
Stratagene mouse T cell 937311	7,134	9,501	0.751
Barstead MPLRB1	6,734	8,907	0.756
Beddington mouse embryonic region	6,424	10,458	0.614
Barstead mouse pooled jejunums MPLRB4	5,994	7,689	0.78
Soares mouse mammary gland NMLMG	5,889	7,249	0.812
Soares mouse placenta 4NbMP 13.514.5	5,398	9,319	0.579
Stratagene mouse macrophage (#937306)	5,107	6,444	0.793
Sugano mouse liver mlia	***4,986***	***6,116***	***0.815***
Life Tech mouse brain	4,828	6,482	0.745
Stratagene mouse diaphragm #937303	4,790	6,316	0.758
Barstead mouse proximal colon MPLRB6	4,402	5,810	0.758
Stratagene mouse testis (#937308)	4,048	5,455	0.742
Stratagene mouse lung 937302	3,659	4,543	0.805
Sugano mouse embryo mewa	***3,434***	***4,582***	***0.749***
Soares mouse uterus NMPu	3,301	4,434	0.744
Stratagene mouse melanoma (#937312)	3,182	4,085	0.779
Stratagene mouse embryonic carcinoma (#937317)	2,923	4,018	0.727
Life Tech mouse embryo 13.5 dpc 10666014	2,876	3,897	0.738
Sugano mouse kidney mkia	***2,657***	***3,336***	***0.796***
Guay-Woodford-Beier mouse kidney day 7	2,631	3,262	0.807
Stratagene mouse kidney (#937315)	2,419	3,479	0.695
Ko mouse embryo 11.5 dpc	2,208	2,664	0.829
Knowles-Solter mouse blastocyst B3	2,203	3,446	0.639
Barstead stromal cell line MPLRB8	1,789	2,087	0.857
Life Tech mouse embryo 8.5 dpc 10664019	1,734	2,367	0.733
Guay-Woodford-Beier mouse kidney day 0	1,728	2,202	0.785
Life Tech mouse embryo 15.5 dpc 10667012	1,425	2,046	0.696
Barstead bowel MPLRB9	1,187	1,558	0.762
Soares mouse hypothalamus NMHy	1,173	1,436	0.817
Stratagene mouse embryonic carcinoma RA (#937318)1161	1,161	1,532	0.758
Life Tech mouse embryo 10.5 dpc 10665016	1,084	1,536	0.706
Soares mouse embryonic stem cell NMES	869	1,144	0.76
Soares mouse urogenital ridge NMUR	572	740	0.773
Knowles-Solter mouse embryonic stem cell	568	761	0.746
Knowles-Solter mouse E6 5d whole embryo	461	768	0.6
Barstead mouse heart MPLRB3	419	735	0.57
Barstead mouse lung MPLRB2	409	1,406	0.291
Knowles-Solter mouse unfertilized egg	338	857	0.394
Barstead mouse testis MPLRB11	306	762	0.402
Knowles-Solter mouse inner cell mass	139	672	0.207
Knowles-Solter mouse 11.5 day limb bud	91	763	0.119
Knowles-Solter mouse 7.5 dpc primitive streak	84	380	0.221
Knowles-Solter mouse 8 cell	79	406	0.195
Barstead mouse spleen MPLRB10	46	738	0.062
Barstead mouse brain MPRB12	25	382	0.065
ESTs submitted to dbEST	344,532	457,778	0.753
ESTs from early developmental stages	124,679	172,067	0.725
ESTs from normalized libraries	**178,500**	**228,859**	**0.78**
ESTs from Sugano libraries	***11,077***	***14,034***	***0.789***

Libraries representing early developmental stages are boxed, normalized libraries are in bold and the Sugano libraries are indicated by italics. The table is sorted by the number of ESTs submitted to dbEST, in descending order. The first column lists the names of the libraries. The second column contains the number of ESTs submitted to dbEST from each library. The third column contains the number of sequences attempted from each library. The final column provides the fraction of sequences submitted to dbEST. Summary statistics for sequences submitted to the database are given at the bottom of the Table.

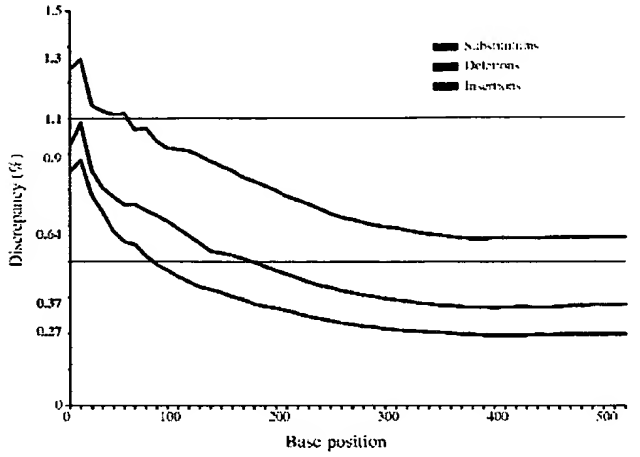

Fig. 1 Sequence discrepancies between the mouse mRNA set and matching ESTs plotted as a function of trimmed sequence length. Discrepancies were categorized by type: substitutions are indicated in red, deletions in blue and insertions in green. Coloured numbers on the ordinate refer to the discrepancy rates at the beginning or end of the trimmed sequence.

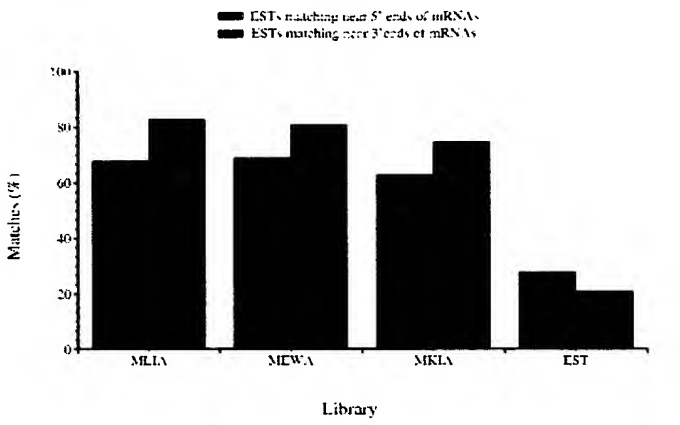

Fig. 2 Sugano libraries are enriched for full-length cDNAs. Shown in red are the percentages of ESTs matching within 50 bp of the 5′ end of an mRNA sequence annotated as full length. Shown in green are the percentages of ESTs matching within 50 bp of the 3′ end of an mRNA sequence annotated as full length. MLIA, MEWA and MKIA denote the Sugano liver, embryo and kidney libraries, respectively. EST indicates data from all other libraries.

We defined the regions of the mRNAs matched by ESTs and found that in 19,920 (28%) cases, the EST match was localized within 50 bp of the 5′ end of the mRNA on the correct strand. These matches may identify full-length or near full-length cDNAs. Late in the project, three oligo-dT–primed libraries potentially enriched for full-length cDNAs (ref. 28) became available. We obtained sequences from the 5′ and 3′ ends of these clones and used these in comparisons with sequences in the mouse mRNA set. Most matches for 5′ ESTs from all three libraries localized within 50 bp of the 5′ end of the matching mRNA (Fig. 2), in contrast to the matches from the larger set of ESTs. The fraction of matching 5′ ESTs may be an underestimate, because some mRNAs in the database probably do not contain complete 5′ UTR. That the Sugano libraries were enriched for full-length sequences and not just for 5′-biased cDNAs was shown by examination of the location of the 3′ matches; most 3′ ESTs matched within 50 bases of the 3′ end of mRNA sequence, (Fig. 2).

Our analysis indicated that, as expected, a large fraction of the ESTs were derived from libraries containing incomplete-length cDNAs. Although this complicated an estimation of the number of genes represented by ESTs, the clustering of related sequences reduced the complexity of the data set. This was performed by comparing ESTs from each library with a larger data set of ESTs. Of 294,835 ESTs analysed, 217,842 were grouped into 20,396 'families', leaving 76,993 'singletons'. We analysed the EST composition of the families, and found 2,109 (10%) contained only ESTs from early-stage libraries. An additional 2,229 (11%) contained ESTs from either early-stage libraries or libraries in which the source material was uncertain. Almost one-third (6,239) of the families contained only ESTs from later-stage libraries. An additional 29% (5,993) of the families contained only ESTs from either later-stage libraries or libraries in which the stage of the source material could not be determined. The remaining 20% (3,799) of the families contained ESTs from early, late and stage-uncertain libraries. The large number of different EST families and singletons indicate a diverse data set; hence, genes expressed at moderate to high levels throughout development are probably well-represented. Accurate enumeration of the number of genes represented requires 3′ ESTs from oligo-dT primed libraries. We have undertaken this activity, and anticipate generating up to 50,000 3′ ESTs in the next six months.

We examined the utility of the mouse ESTs in inter-species gene identification. Using stringent criteria, we found that 81% of the sequences in a non-redundant human mRNA database (G. Schuler, pers. comm.) were matched by at least one mouse EST. In another assay, both human and mouse ESTs were searched against 76.7 million base pairs of human genomic sequence generated by the Human Genome Project. Although 3.1% (2.38 Mb) of this sequence was matched by either a human or mouse EST, more than 0.47% (360,000 bp) were matched only by mouse ESTs. The mouse ESTs thus represent a rich new source of conserved sequences that can be exploited for gene-finding purposes. The utility of ESTs are not limited in this regard in mammals; a comparison of translated mouse ESTs with a set of 1,517 proteins conserved between yeast and *Caenorhabditis elegans* revealed that more than 92% of conserved proteins were matched by a mouse sequence. The mouse ESTs thus offer the possibility of identifying similar sequences from organisms as distantly related as fungi and nematodes, facilitating the use of these powerful experimental systems in exploring the functions of potential homologues.

The ESTs described here provide a broad overview of genes expressed throughout the development of the laboratory mouse, and lend themselves to a variety of applications. They provide an enormous number of entry points into lines of investigation that can be undertaken in parallel. By providing rapid access to many mouse genes well in advance of large quantities of mouse genome sequence, the ESTs have enhanced the value of the mouse as a model for biology. As increasing amounts of genome sequence become available, ESTs will provide an indispensable tool for interpreting it. The first step in identifying a mouse homologue can now be taken using a computer.

Methods

DNA purification and sequencing. Bacterial clones were plated, colonies picked robotically and glycerol stocks constructed in 384-well format. Clones were grown, DNA prepared and sequencing performed as described[12] (M.M. *et al.*, manuscript submitted). Estimates of cDNA size were not generated. As with our human EST project[12], clones were arrayed and distributed by the Lawrence Livermore National Laboratory-based I.M.A.G.E. consortium[29] to commercial distributors (see http://www-bio.llnl.gov/bbrp/image/image.html for details) to provide the scientific community with access to the clones.

Computational analysis. Our analysis was performed on a set of 295,053 mouse ESTs available as of 1 April 1998. Of these, 116,220 (39%) were from libraries prepared from embryonic tissue, 172,714 (59%) were from libraries prepared from later-stage tissues and 5,901 (2%) were from sources difficult to classify. Before cluster analysis, sequence repeats were masked using 'repeatmasker' with the –m option (A. Smit, pers. comm.). Clustering was performed using BLASTN2 (http://blast.wustl.edu, W. Gish, pers. comm.; S=300, gapS2=150, M=5, N=–11, R=11, Q=11, filter seg) to compare all ESTs with each other. All similarities with *P*-values better than 10^{-99} were evaluated to ensure they met the 97% identity and match length (at least 50 bp) cutoffs. Only those ESTs with matches consistent with their membership in a single cluster were considered. BLASTN2 (S=300, gapS2=150, M=5, N=–11, Q=11, R=11, B=5,000, V=5, filter seg) was used to compare human ESTs with human mRNAs (6,444 sequences) and mouse ESTs with mouse mRNAs (3,640 sequences). Before performing the comparisons, mammalian repeats found in the sequences were masked using 'repeatmasker' (A. Smit, pers. comm.). To compare human ESTs with mouse mRNAs and mouse ESTs with human mRNAs, S was relaxed to 170 and N to –5. Cutoff *P*-value scores were 10^{-99} or 10^{-49} for same-species or cross-species matches, respectively. Genomic sequences (1,569) totaling 76.7 Mb were extracted from the High-Throughput-Genome Sequence division (Phase 3 finished) of GenBank. Repeats were masked in 'default' mode to mask primate-specific and mammalian-wide repeats and in '–m' mode to mask mouse- and other rodent-specific repetitive elements. Mouse ESTs, likewise masked for rodent and mammalian-wide repeats, and human ESTs, masked for human repeats, were compared with the human genomic sequence using BLASTN2 (S=170, gapS2=150, M=5, Q=11, R=11, filter seg, N=–11 for the human ESTs and N=–5 for the mouse ESTs). As above, cutoff *P*-value scores were 10^{-99} or 10^{-49} for same-species or cross-species matches, respectively.

A complete set of 6,221 yeast proteins was compared with 13,747 worm proteins (Wormpep13; ref. 30) using BLASTP2 (http://blast.wustl.edu; W. Gish, pers. comm.) with the parameters (V=0, H=0, –hspmax=100,000, M=BLOSUM62, filter seg). The program BLASTX2 (V=0, H=0, –hspmax=100,000, M=BLOSUM62) was then used to compare each of the mouse ESTs with the set of 1,517 proteins conserved between *C. elegans* and yeast. In these experiments, a *P*-value cutoff score of 10^{-9} was considered indicative of a match.

Acknowledgements

We thank all investigators who have donated libraries for sequencing; S. Tilghman for scientific guidance; S. Chissoe and S. Gorski for comments on the manuscript and useful discussion; G. Schuler, C. Tolstoshev and others at NCBI for assistance with databases; and the staff at Washington University Genome Center for technical support. Work by C.P. and G.L. was supported by the U.S. DOE under contract W-7405-Eng-48 to LLNL. Work at Washington University was funded by a grant from Howard Hughes Medical Institute.

Received 17 November; accepted 21 December 1998.

1. Brown, S.D.M. & Peters, J. Combining mutagenesis and genomics in the mouse—closing the phenotype gap. *Trends Genet.* **12**, 433–435 (1996).
2. Zambrowicz, B.P. *et al.* Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature* **392**, 608–611 (1998).
3. Hicks, G.G. *et al.* Functional genomics in mice by tagged sequence mutagenesis. *Nature Genet.* **16**, 338–344 (1997).
4. Milner, R.J. & Sutcliffe, J.G. Gene expression in rat brain. *Nucleic Acids Res.* **11**, 5497–5520 (1983).
5. Putney, S.D., Herligh, W.D. & Schimmel, P. A new troponin T and cDNA clones for 13 different muscle proteins, found by shotgun sequencing. *Nature* **302**, 718–721 (1983).
6. Adams, M.D. *et al.* Complementary DNA sequencing: expressed sequence tags and the human genome project. *Science* **252**, 1651–1656 (1991).
7. Adams, M.D. *et al.* Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* **377**, 3–17 (1995).
8. McCombie, W.R. *et al. Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease gene homologues. *Nature Genet.* **1**, 124–131 (1992).
9. Waterston, R.H. *et al.* A survey of expressed genes in *C. elegans*. *Nature Genet.* **1**, 114–123 (1992).
10. Sasaki, T. *et al.* Toward cataloguing all rice genes: large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library. *Plant J.* **6**, 615–624 (1994).
11. Houlgatte, R. *et al.* The GenExpress index: a resource for gene discovery and the genic map of the human genome. *Genome Res.* **5**, 272–304 (1995).
12. Hillier, L. *et al.* Generation and analysis of 280,000 human expressed sequence tags. *Genome Res.* **6**, 807–828 (1996).
13. Yamamoto, K. & Sasaki, T. Large-scale EST sequencing in rice. *Plant Mol. Biol.* **35**, 135–144 (1997).
14. Nelson, P.S. *et al.* An expressed-sequence-tag database of the human prostate: sequence analysis of 1168 clones. *Genomics* **47**, 12–25 (1998).
15. Ajioka, J.W. *et al.* Gene discovery by EST sequencing in *Toxoplasma gondii* reveals sequences restricted to the Apicomplexa. *Genome Res.* **8**, 18–28 (1998).
16. Sasaki, N. *et al.* Characterization of gene expression in mouse blastocyst using single-pass sequencing of 3995 clones. *Genomics* **49**, 167–179 (1998).
17. Sutherland, H.F., Kim, U.J. & Scambler, P.J. Cloning and comparative mapping of the DiGeorge syndrome critical region in the mouse. *Genomics* **52**, 37–43 (1998).
18. Makalowski, W. & Boguski, M.S. Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. *Proc. Natl Acad. Sci. USA* **95**, 9407–9412 (1998).
19. Makalowski, W., Zhang, J. & Boguski, M.S. Comparative analysis of 1,196 orthologous mouse and human full-length mRNA and protein sequences. *Genome Res.* **6**, 846–857 (1996).
20. Scharf, J.M. *et al.* Identification of a candidate modifying gene for spinal muscular atrophy by comparative genomics. *Nature Genet.* **20**, 83–86 (1998).
21. Bailey, L.C. Jr, Searls, D.B. & Overton, G.C. Analysis of EST-driven gene annotation in human genomic sequence. *Genome Res.* **8**, 362–376 (1998).
22. Jiang, J. & Jacob, H.J. EbEST: an automated tool using expressed sequence tags to delineate gene structure. *Genome Res.* **8**, 268–275 (1998).
23. Schena, M. *et al.* Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol.* **16**, 301–306 (1998).
24. Schuler, G.D. *et al.* A gene map of the human genome. *Science* **274**, 540–546 (1996).
25. Bonaldo, M.F., Lennon, G. & Soares, M.B. Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res.* **6**, 791–806 (1996).
26. Ewing, B., Hillier, L., Wendl, M. & Green, P. Basecalling of automated sequencer traces using PHRED I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
27. Ewing, B. & Green, P. Basecalling of automated sequencer traces using PHRED II. Error probabilities. *Genome Res.* **8**,186–194 (1998).
28. Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. & Sugano, S. Construction and characterization of a full length-enriched and a 5'-end enriched cDNA library. *Gene* **200**, 149–156 (1997).
29. Lennon, G., Auffray, C., Polymeropoulos, M. & Soares, M.B. The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics* **33**, 151–152 (1996).
30. Sonnhammer, E.L. & Durbin, R. Analysis of protein domain families in Caenorhabditis elegans. *Genomics* **46**, 200–216 (1997).